DOCUMENT RESUME

ED 209 657 CS 006 374

AUTHOR Altwerger, Bess; Goodman, Kenneth S.
TITLE Studying Text Difficulty through Miscue Analysis. Program in Language and Literacy. Occasional Paper Number 3.
INSTITUTION Arizona Univ., Tucson. Coll. of Education.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.; Office of Education (DHEW), Washington, D.C.
PUB DATE Jun 81
NOTE 30p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; Cross Cultural Studies; Elementary Education; Error Analysis (Language); Interaction Process Analysis; Interference (Language); *Miscue Analysis; *Reading Comprehension; *Reading Difficulties; Reading Processes; *Reading Research; Structural Analysis (Linguistics); *Syntax
IDENTIFIERS *Prose Learning

ABSTRACT

As part of a larger study of the oral reading of elementary school students representing eight linguistic populations in the United States, a study was conducted to discover why readers make the same miscues at the same point in a text and to discover factors in the text that contribute to this phenomenon. Subjects were second, fourth, and sixth grade students who were Navajo, Hawaiian Samoan, Arab, and Texas Spanish second language speakers, as well as downeast Maine, Appalachian white, Mississippi rural black, and Hawaiian-pidgin dialect speakers. They were instructed to read aloud whole stories of considerable length and to recall all they could remember about the stories. Sentences that generated the highest rates of miscues per word per reader were then analyzed for aspects that contributed to those rates. The analysis confirmed that syntactic complexity was not the only contributor to miscues. Other factors causing miscues were (1) lack of relevant prior knowledge, (2) unfamiliar or unusual use of terminology, (3) weak syntax, (4) unpredictable simple structures, (5) unusual stylized syntax, (6) complex syntax, and (7) combinations of the above. The findings suggest that text difficulty cannot be understood completely without some investigation of the interaction between readers and the text, and that miscue analysis can provide data that reveal such interaction. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209657

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Studying Text Difficulty Through Miscue Analysis

Bess Altwerger
University of New Mexico

Kenneth S. Goodman
University of Arizona

A Research Paper

June, 1981 No. 3

Occasional Papers
Program in Language and Literacy
Arizona Center for Research and Development
College of Education
University of Arizona

Co-directors:
Kenneth S. Goodman
Yetta M. Goodman
402 Education, Bldg. 69
University of Arizona
Tucson, AZ 85721

Portions of the data in this study were generated through studies supported by the U.S. Office of Education and The National Institute of Education. No endorsement of the statements herein is implied.

CS206374

Abstract

Sentences which generated the highest rate of miscues per word per reader were analyzed for aspects which contributed to the high miscue rates. Correlations between miscue rate for all sentences in each of three stories and the Schmidt-Kittel Linguistic Complexity Ratio were also obtained. These correlations for each story were significant but moderate (.27, .23, .38 respectively).

Analysis of the sentences confirmed that syntactic complexity itself was not the only contributor to miscues. These aspects emerged: 1) Lack of relevant prior context; 2) Unfamiliar or unusual use of terminology: 3) Weak syntax; 4) Unpredictable simple structures; 5) Unusual stylized syntax; 6) Complex syntax; 7) Combinations of all.

The study was part of a larger study of second, fourth, and sixth graders in eight populations of American readers with different language backgrounds.

The authors conclude that text difficulty can not be truly understood without investigating the interaction between readers and the text. Miscue analysis provides data that reveal that interaction.

STUDYING TEXT DIFFICULTY

THROUGH MISCUE ANALYSIS

The focus of miscue research has been on what we can learn about the reading process through the analysis of readers' miscues. This research has provided us with important insights into the kinds of information and strategies readers utilize in constructing meaning from print. In this paper, however, we make a 90° turn and look at what we can learn about text difficulty through the miscues our subjects have made. To do so, we chose sentences which had the highest relative frequency of miscues from three standard stories. Our concern was with understanding why many readers will make miscues at the same point in a text, and to discover factors in the text which contribute to this phenomenon.

This study on text difficulty is part of a larger federally funded miscue research study (Goodman & Goodman, 1978), which analyzed the oral reading of second, fourth, and sixth graders representing eight linguistic populations. These populations are Navajo, Hawaiian Samoan, Arab and Texas Spanish second language speakers, as well as Downeast Maine, Appalachian White, Mississippi Rural Black and Hawaiian Pidgin dialect speakers.

As in all miscue research, subjects were instructed to read aloud whole stories of considerable length and to later retell all they could remember about the stories. At each grade level, subjects read one "standard" story

---

The research reported herein was supported in part by the National Institute of Education, Department of Health, Education and Welfare. However, the opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement by NIE should be inferred.

chosen from the Betts Basic Readers (1963)*. The oral reading and retelling of the stories were taped and later analyzed.

Miscues are points in oral reading where the observed response of the reader does not match the expected response. Miscues are analyzed by means of the Goodman Taxonomy, which compares the observed response to the expected response on variables which include graphic and phonemic proximity, syntactic and semantic acceptability and change, morphemic involvement, intonation (see Allen and Watson, 1976 for complete taxonomy).

Miscue Frequency Measures

Several quantitative measures of miscue frequency have been used to gain insight into where and why miscues cluster. For each sentence of the stories used in this study, the following was computed:

1. MISCS - the total number of miscues produced on each sentence.
2. MPWD - Miscues per word. This measure allows for a comparative analysis of miscue frequency for sentences of varying word lengths within a story.
3. MPWPR - Miscues per word per reader. This would be the most useful figure for comparison across studies with different numbers of subjects.

Linguistic Complexity

In addition to the above calculations, the syntactic complexity of each sentence was analyzed through the use of the "Schmidt-Kittel Linguistic Complexity Scale."** This scale is weighted to include points for Operations,

* In the larger study each language group also read a "culturally relevant" story but those readings are not involved in this sub-study.

** We are indebted to Eunice Schmidt, Seattle Pacific University for performing this analysis on the three stories.

"the term given to the manipulations or movements occurring in measuring syntactic complexity to operationalize the process numerically" (Schmidt, Kittel). The number of total operations per sentence is then divided by the number of words per sentence, thereby yielding the Linguistic Complexity Ratio. The complexity scale reflects such structural elements as elaborated phrases and clauses, unusual word order (preposing or postposing), unusual and varied vocabulary, anaphoric structures, and the extent to which surface structure implies the deep structure. Though it includes some semantic factors, it primarily focuses on syntactic complexity.*

Operations and Miscue Frequency

Pearson correlation coefficients were computed to assess the relationship among the following variables:

sentence length in words (WORDS)

number of miscues per sentence (MISCS)

miscues per word (MPWD)

miscues per word per reader (MPWPR)

operations per sentence (OPERS)

operations per word, or the Syntactic Complexity Ratio (OPPWD)

Table 1 presents the significant correlations found between these variables within each of the three standard stories read by the subjects.

*We chose this measure because of its focus on syntactic complexity. We make no claim for this being a definitive measure of syntactic complexity. It is one measure, based in sound linguistics. As such it serves our purpose which is to consider the extent to which complexity itself is the cause of high miscue rates.

Table 1

Complexity and Miscue Frequency

	Story #44*	Story #51**	Story #53***
MISCS X WORDS	r = .6224 s = .001	r = .8091 s = .001	r = .6923 s = .001
OPERS X WORDS	r = .9304 s = .001	r = .9642 s = .001	r = .9464 s = .001
MISCS X OPERS	r = .6720 s = .001	r = .8141 s = .001	r = .7614 s = .001
OPPWD X MPWD	r = .2673 s = .006	r = .2264 s = .003	r = .3756 s = .001
OPPWD X MPWPR	r = .2672 s = .006	r = .2311 s = .002	r = .3798 s = .001
WORDS X MPWPR	NS	NS	NS

* Kitten Jones
** Freddie Miller, Scientist
*** My Brother Is A Genius

A very high positive correlation, significant at the .001 level, exists between the number of operations (OPERS) and sentence length (WORDS). The longer the sentence, the greater the linguistic complexity, according to the Schmidt-Kittel computation. Since a moderate correlation was also found between total number of miscues (MISCS) and sentence length (WORDS), it is not surprising that a slightly higher significant relationship also exists between operations (OPERS) and miscue frequency (MISCS). However, when frequency of operations (OPPWD) and miscues (MPWD) are adjusted for sentence length, the positive relationship between operations and miscues is significant but modest (.23 to .38). This indicates that the relationship between miscue frequency (MISCS) and operations (OPERS) is more a result of sentence length than the complexity ratio itself. There is no significant correlation between miscues per word per reader (MPWPR) and sentence length (WORDS).

Sentences Producing High Number of MPWPR

Table 2 presents the sentences selected from each story which resulted in the highest rate of miscues per word per reader for that story. This number, as well as the word length and operation ratio for each sentence, has been listed.

Table 2

Sentences with Highest Miscue Rates

Story Number	Sentence Number	Sentence	WORDS	OPPWD	MPWPR
53*	8	"Philosophical" I yelled.	3	5.00	.490
53	14	"Philosophical" I shouted.	3	4.33	.391
53	26	Sinewy: stringy, strong, or powerful.	5	6.00	.425
53	211	"Sleigh, snow, soak, society, soften, soldier, sorrowful, soap, stormy, soak, survive.	11	6.72	.477
53	167	There were glaring spot-lights and floodlights and cables rigged up every-where.	11	4.81	.369
53	118	"Say da", Mr. Barnaby chuckled.	5	3.60	.319
Story Means				3.76	.123
51**	5	"You've wrecked that doll!" she exclaimed.	6	5.50	.275
51	66	Mr. Miller sighed.	3	2.33	.302
51	22	After the cut in his allow-ance, Freddie's chemistry experiments narrowed to those safely outlined in a library book.	18	4.44	.240
51	73	"In the hall closet" came Elizabeth's tearful reply.	8	4.87	.305
51	80	His sister's cries grew louder.	5	4.60	.275
51	134	Such quick thinking	3	5.66	.302
Story Means				3.79	.113

* My Brother Is A Genius
** Freddie Miller, Scientist

Table 2

Sentences with Highest Miscue Rates

(Cont'd)

Story Number	Sentence Number	Sentence	WORDS	OPPWD	MPWPR
44***	15	There are baseballs, bats, marionette dolls, and big balloons" said Penny.	11	4.54	.320
44	16	"Marionette dolls" exclaimed Sue.	4	3.75	.420
44	48	He printed them upstairs in his dark room.	7	4.28	.330
44	54	"How clear it is!"	4	4.25	.340
44	76	The judges laughed.	3	2.33	.360
Story Means				3.40	.151

*** Kitten Jones

While the majority of OPPW's for each sentence are above the story means, a number of sentences do fall below the mean. Both the OPPW's and the sentence lengths within each story vary considerably. The mean ratio of MPWPR for the three stories are similar. However, in comparing the sentences within each story, we find that the selected sentences in Story 51 do not produce as high a rate of MPWPR as those in the other two stories. In fact, Story 53 had several more sentences that produced MPWPR that compare to the highest on Story 51. We can only conclude that the means do not reveal the full picture and that stylistic differences may, in fact, be involved. Data in the larger study indicates that Story 53 is not a harder task for sixth graders than Story 51 is for fourth graders.

Results from the data presented in both Tables 1 and 2 indicate that miscue frequency is not simply a function of either sentence length or linguistic complexity (as measured by the Schmidt-Kittel Scale). For instance, five of the sentences with highest MPWPR consist of only three words. This is important to note, as sentence length is often a main consideration in assessing readability, due in part to the relationship believed to exist between sentence length and linguistic complexity. The existence of this relationship has been supported by our data (see Table 1). However, while linguistic complexity does seem to be a factor in miscue frequency for some sentences, it is not, alone, a reliable predictor of difficulty as shown by miscue frequency.

Miscue frequency cannot be explained solely by factors related to the written language encoded by the author. This is consistent with our theoretical base in that reading is viewed as an <u>interaction</u> between the author and

the reader; a communication process. Readers are active participants in this process, who utilize their knowledge of language, their past experiences, background and concepts in order to make predictions about the meaning and structure of the text. It follows then, that the closer the author's experiences, language and concepts are to those of the reader, the more effective the communication. Miscues will occur when certain lexical items, syntactic structures, concepts or events introduced in the story are unexpected, unfamiliar or in some other way difficult for the reader to predict. Therefore, in order to fully understand the factors contributing to miscue frequency, we must consider the written text in relation to, and not separate from, the reading process itself. We must analyze what makes these sentences with the highest rate of MPWPR difficult for readers of varying linguistic and cultural backgrounds to predict.

Lack of Contextual Support

When the language or concepts within a story are unfamiliar to the reader, redundancy or strong contextual support provides additional information that the reader can use to formulate predictions.

For several sentences, a careful analysis of the preceding portion of the stories and the miscues produced indicates that there are none or few contextual cues which the reader may utilize in order to predict what is to follow. It was also noted that these sentences are relatively simple structures, each consisting of three words. In Story 53, sentences 8 and 14 both produce high MPWPR. The sentences are:

Sentence 8 - "Philosophical!" I yelled.

Sentence 14 - "Philosophical!" I shouted.

Both these sentences share the same syntactic structure and contain the word "Philosophical". Directly preceding sentence #8, the reader is informed that

the main character will be choosing, at random, a word to read from the dictionary. Therefore, the only cues the reader has available are the graphophonic cues. The grammatical structure offers little support, in that any form class of words could fit as well into the sentence slot which "philosophical" fills. The form class of the word would also be of little consequence to the meaning of the story in general. Thus, the miscues produced consist either of non-words with high graphic and phonemic similarity to the ER, or omissions. Sentence 14 follows a "definition" of philosophical: showing calmness and courage in the face of ill fortune. It is highly questionable that this can be regarded as a definition of philosophical at all. The high number of miscues for sentence 14 indicate that for the children reading this story, the definition offers no further cues.

Sentence 76 - "The judges laughed" - in Story 44 is another example of those high MPWPR sentences for which there are few supporting contextual cues. This sentence has an OPPW ratio of 2.33, falling below the story average of 3.40. The majority of miscues for this sentence involve the word "judges". In analyzing the preceding story line, it becomes evident that there is a sudden change in setting, time, sequence, and characters without a clear transition by the author. It must be inferred by the reader that there is a shift into a future time period, that a contest judging is now in progress and that there are judges involved in the scenario. Furthermore, based on children's experiences with courtroom scenes on TV, etc., it would be logical to assume that one judge would be involved in the contest. In fact, most of the miscues are substitutions of a singular form of the plural form of judge. Other miscues include non-word substitutions, and syntactically and semantically unacceptable substitutions. Thus, a lack of contextual support for predicting

particular lexical items, structures or events in a story can, in and of itself and in conjunction with other factors (discussed later), be a source of high MPWPR.

Unfamiliar or Unusual Lexical Items

In the examples above, one might argue that "hard" words caused the difficulties. One must consider, however, when such difficult lexical items cause problems. Those we have cited had little contextual support.

Several sentences generating high MPWPR do include a lexical item which accounts for a great many of the miscues for those sentences.

A lexical item can be difficult for various reasons, ranging from position in a particular syntactic structure to the frequency with which it occurs in the reader's linguistic environment. A lexical item may rarely occur in a reader's environment if it is a technical term or part of a specialized vocabulary for a particular field of study. Often, one lexical item can have several general meanings as well as a technical meaning, and may be interpreted in a variety of ways, depending upon the reader's knowledge, background and concepts. The problem is much more complicated than simply knowing or not knowing the word.

In Story 44, sentence [redacted] is "There are baseballs, bats, marionette dolls, and big balloons" said Penny. The lexical item, marionette, generates many miscues. This word also occurs in sentence 16, "Marionette dolls!" exclaimed Sue, which again generates a high number of MPWPR.

The word marionette is a specialized term for a particular kind of puppet; one operated by the manipulation of strings. The word puppet is probably a more familiar and all-encompassing term used by those without a specialized

knowledge of this art form. It is interesting, however, to note that the miscues involving marionette in sentence 15 are qualitatively different from those produced for the same word in sentence 16.

Substitutions for marionette in sentence 15 are generally semantically and syntactically acceptable such as more dolls, other dolls, Mattel doll, marching dolls. The same readers, however, move to either non-word substitutions such as $monching dolls, $mahale dolls, or omissions for marionette dolls in sentence 16. This change in miscue quality may be due to the fact that sentence 15 provides a conceptual and syntactic framework which the reader can utilize for prediction, while sentence 16 does not. One reader made particular use of the conceptual framework of sentence 15 to produce mitts as a substitute for marionette, which follows baseballs and bats.

Other miscues in sentence 15 include such substitutions as basketballs for baseballs and the treatment of ...baseballs, bats... as one unit (a very common unit) -- baseball bats. Other miscues in sentence 16 generally involve exclaimed, a term rarely, if ever, used in oral language. Explained is a frequent substitution.

Sentence 48 in Story 44 - He printed them upstairs in his darkroom - represents an example of a sentence which utilizes common words with technical meanings. In this case, a knowledge of photography, as well as a conceptual framework for film development and photographic processing, is a prerequisite to the interpretation that the author most likely had in mind. This more technical interpretation of the sentence is, however, made even less predictable due to the text directly preceding this sentence: Mr. Jones finished the pictures himself. Note that the word picture, rather than photograph, is used

here and throughout the story. Although there is mention of camera and the taking of pictures throughout the story, the concept of finishing the pictures in terms of photography may be quite alien to the reader. Many miscues consisted of substituting the word painted for printed, indicating that the reader conceptualized finishing the picture, in this context, in terms of their own experiences of finishing pictures: with paint or crayons. The high graphic similarity between print and paint would support this prediction. As would be expected, intonation indicates that darkroom, here referring to the room in which developing and finishing takes place, was frequently processed by the readers as two words - dark room, consisting of an adjective and noun. Clearly, the readers are constructing a meaning for this sentence which is appropriate to their knowledge, concepts and experiences. In this case, however, the author presupposes knowledge and experiences that do not coincide with those of the readers.

Syntax

The significance of syntax has been considered in the development of some readability formulas. Those such as the Dawkins, Botel and Granowsky Syntactic Complexity Formula, (1973) are based on the assumption that in regard to syntax, the more complex the syntax (the number of deletions, postposing, fronting, etc.) the more difficult the readability. Although this does seem to be a factor in causing high MPWPR in some cases, syntactic factors other than complexity may contribute to the miscue frequency. Analysis of the sentences generating high MPWPR in this study reveals several such syntactic features.

Weak Syntactic Structure

To get to meaning readers predict the syntactic structure based on their knowledge of the language. The process of constructing meaning also requires

using syntactic patterns to confirm and correct prior predictions. When the syntactic structure is not easily predicted or recognized or no syntactic structure is available at all, readers must rely more heavily on other cuing systems such as the graphophonic.

Sentence 211 in Story 53 is a good example of such a case. The "sentence" is simply a list of words read in alphabetical order from a dictionary: Sleigh, snow, soak, society, soften, soldier, sorrowful, soap, stormy, stroke, survive... There is no syntactic structure at all: each word is a separate entity. There is no syntactic or semantic context, so only word identification strategies are utilized by the reader. The words in this sentence are completely random with the limitation that they begin with the initial consonant s. Unlike sentence 15 in Story 44 - "There are baseballs, bats, marionette dolls, and big balloons" said Penny - there is not even a conceptual framework within which the items listed fall. There is neither a conceptual nor syntactic relationship between any of the words listed in this sentence.

The miscues on sentence 211 were generally substitutions of non-words and real words, most of which begin with the initial consonant s. Exceptions to this are substitutions such as often for soften and drove for stroke. The sentence was generally read with the intonation that one might expect to use when reading a list of words. However, the high number of MPWPR (.477 - the second highest for all sentences in the study) indicates that this type of sentence, which lacks many of the cuing systems normally present in written language, is particularly difficult to read. The cue systems of language must support each other to aid the reader.

Predictability and Syntactic Structures

Readers must predict syntactic structures well before they have read all the words in them.

In many structures, the first word of the sentence provides reliable and important information about the total sentence and is a good source of prediction for readers. For instance, if why is the first word of a sentence, readers take little risk in assuming that the structure will be an interrogative. Based on readers' knowledge of the structure of interrogatives in English, they may also predict other more specific features of the sentence; for example, that the word following why will probably be either a modal, have or be. Likewise, in sentence 54 of Story 44, How clear it is, readers who use the first word to predict a question will most likely expect the features of an interrogative sentence. How, of course, often serves the function of question marker accompanied by an inversion of the subject and auxiliary. However, this sentence turns out not to be an interrogative but an active, declarative exclamation of a rather peculiar type. (Compare: It is so clear.) Thus, as we would expect, many of the miscues involve either a reversal of the order of it is, resulting in is it, and thereby following through the prediction of an interrogative, or omissions of it, followed by a regression to correct after is. In addition, many readers substitute other adjectives such as clean and clever for clear, resulting in syntactically acceptable structures.

These miscues indicate that readers are using their knowledge of the structure of English sentences to make logical predictions concerning the syntactic features of the sentences they read.

Stylized Syntax and Metaphor

The manipulation of syntactic form is a common means by which authors can create and express their own literary style. While the resulting stylized structures may be aesthetically pleasing to the author and the readers, conceptual and linguistic predictability is often sacrificed in the process. To achieve novelty, we sacrifice predictability.

Several sentences in this study which generated high MPWPR fall within this category. They are generally literary structures which may be difficult for children to predict. For instance, several contain metaphors which violate selection restrictions by combining inanimate nouns with verbs which normally require animate subjects, such as the verb came with the noun reply. Others contain intransitive verbs such as chuckle, used in a transitive sense as a dialogue carrier. Children's miscues are evidence of their attempts to construct meaningful syntactic structures consistent with the story content.

Sentence 73 of Story 51 - "In the hall closet" came Elizabeth's tearful reply - contains several literary features which make this sentence conceptually and linguistically hard to predict and comprehend. The verb came, for instance, serves two functions in this sentence: 1) Elizabeth replied by saying "(I am) In the hall closet"; 2) The reply came from the hall closet. In addition, the use of tearful to modify the noun reply is, of course, a metaphoric device: Literally, the "reply was full of tears", but meaning she replied tearfully.

The miscues for this sentence indicate the readers' often successful efforts in breaking through the surface structure to discover the deep structure and the logical relationships underlying the lexical items. For instance, several mis-

cues involve a substitution at the word level,(and insertion of a suffix at the morphemic level) of tearfully for tearful. These miscues accurately reflect the deep structure relationships of Elizabeth replied tearfully, in which tearfully is an adverb modifying Elizabeth's act of replying. These miscues result in structures such as came Elizabeth's tearfully replied and came Elizabeth tearfully reply.

Other miscues for this sentence involve the substitution of Elizabeth for Elizabeth's, thereby making Elziabeth the subject of came, a more predictable logical subject for the verb came than reply.

Sentence 80 of Story 51 is another example of how stylistic features can cause complexity. The majority of miscues for the sentence, His sister's cries grew louder, involve the possessive sister's cries in relation to the verb grew. It's important to note that the word cries can be a verb in the sense of weeping or it can be either a verb or noun in the sense of calling out. This sentence contains the latter sense of cry as a plural noun. However, in the previous context the reader is told that Elizabeth is indeed weeping, thus making the weeping of cry highly predictable. The miscues clearly indicate that this is true. A great many miscues delete the possessive 's from sister's, transforming his sister's cries into his sister cries or cried, in which sister is the subject of the verb cries or cried. Thus, cries takes on the sense of weeping, and is in accord with the story line. Several readers then omit grew which would conflict with his sister cries, thus producing his sister cries (or cried) louder. These miscues render a non-metaphoric interpretation of the sentence and eliminate the tension caused by the violation of selection restrictions for cries grew. Others regress to correct at this point, or leave the structure as a syntactically and semantically unacceptable sentence.

Sentence 118 in Story 53 is "Say da", Mr. Barnaby chuckled. It exemplifies a widely used stylistic feature found in children's literature. Perhaps, in attempting to avoid repetitive use of "said", "answered" or "replied", many authors use such constructions as laughed Bob, cried Mary, Jim giggled, or, in this sentence, Mr. Barnaby chuckled. The word chuckled, if ever encountered in oral language, would probably be used as an intransitive verb. In this sentence, however, it is used as a transitive verb with "say da" as its object. In addition to this, the quote itself "say da" is unusual in the sense that a non-word is used as object of an imperative verb with the subject deleted so that it must be inferred by the reader.

The miscues for this sentence indicate that many readers processed it as an interjection rather than an imperative, inserting a comma after say, resulting in say, da with intonation similar to Say, John, how is Mary? Several readers also substituted a real word, either dad or daddy for da, a logical prediction based on what is normally found in written language. Another observation based on the miscues for this sentence is that the one sentence was processed by many readers as two separate sentences, in which Mr. Barnaby has not uttered the command Say da. In other words, the intonational pattern suggests that a period was inserted to produce Say da. Mr. Barnaby chuckled. Say da, in this case, is not the object of chuckled, but rather, chuckled is interpreted as an intransitive verb.

It seems clear that the authors' styles have contributed to linguistic and conceptual complexity as reflected in the readers' miscues. In each case, the readers attempt to eliminate the syntactic or semantic violations the author employs as stylistic devices.

Complex Syntactic Structure

Sometimes as our correlations indicated, miscues do reflect sheer syntactic complexity in the sense mentioned earlier in this section; that is, having undergone various transformations such as preposing, elipses, fronting, relative clause deletion, etc. Sentence 22 of Story 51 is After the cut in his allowance, Freddie's chemistry experiments narrowed to those safely outlined in a library book. It contains several complex features which are reflected in the miscues of the readers.

The sentence begins with a left branching dependent clause with a complicated surface structure with the predicate deleted (the cut in his allowance was made). The pronoun his within this clause is co-referential with the proper noun Freddie, which occurs as the subject noun in the following independent clause. The pronoun those, which occurs in the prepositional phrase following the main clause verb phrase, refers ambiguously to either the types or numbers of chemistry experiments or the actual chemistry experiments themselves. Following those is a reduced relative of the underlying structure those (which were) safely...with which were deleted. The use of the term safely outlined is misleading in that it actually refers to safe experiments which were outlined. This entire clause is in the passive with the agent deleted.

The points at which miscues cluster in this sentence indicate which features might be most complex or most syntactically ambiguous. Many of the miscues involve the first clause of the sentence. The noun phrase the cut is changed frequently to either he cut or they cut, resulting in a subject and verb in place of the deleted one. The cut in the text is a nominalization of a verb phrase from someone cut his allowance.

His allowance is replaced frequently by the allowance, which, of course, loses the co-referentiality of his with Freddie. It is important to note that a causal relationship between Freddie's previous experiments discussed in the story and the cut in his allowance by Freddie's mother as punishment must be inferred simply from the phrase after the cut in his allowance. The miscues of they cut or he cut for the cut indicate that the reader has not inferred that Freddie's mother is the one responsible for cutting Freddie's allowance. The miscues of the allowance for his allowance suggests that the readers may not be aware of whose allowance is being cut. Thus, this prepositional phrase, with a pro-form whose reference is not immediately discernable, is quite complex and inexplicit. In addition, the causal relationship which underlies the meaning of this sentence is not explicitly and clearly stated.

The subject noun phrase in the main clause begins with the possessive form of Freddie's. Many readers, expecting the subject noun to be the first word in the phrase, substitute Freddie for Freddie's, and then expect chemistry to be a verb.

In the reduced relative clause preceded by those, many readers turn the structure into those safety... in which those is a determiner and safety is an adjective. Either the reduced relative clause is not assigned by the reader or the complexity mentioned earlier concerning safely outlined has contributed to the construction of these miscues.

The analysis of this sentence seems to indicate that the syntactic features which are often considered linguistically complex as a result of various transformations, can, in fact, generate a large number of miscues. The miscues provide us with insights into the ways in which these syntactic features interact with readers' predictions and expectations, and the extent to which relationships in the story are clearly expressed by the surface structure representations.

Combination of Factors

This category includes those sentences in which combinations of the factors previously outlined seem to contribute to the high miscue frequency. In other words, these sentences can have unusual lexical items, a lack of contextual support, in addition to various other features.

Sentence 26 of Story 53: Sinewy: stringy, strong or powerful is an example of this type of sentence. It is a definition of a word which was chosen at random from a dictionary to be read aloud by the main character. There is no prior information provided that would be helpful to the reader in predicting that this particular word would be read. The reader does, however, have contextual clues that suggest that a dictionary definition will be read aloud by the character. Sinewy is probably a low frequency word in the children's linguistic environments, and therefore unpredictable. The syntactic structure is rather weak in that it lacks an overt basic sentence order of subject-verb-object. However, the punctuation (the colon) supplies a structure in the sentence so that it serves as a verb marker. The sentence can be paraphrased as Sinewy is defined as...or Sinewy means... The colon makes these interpretations possible, but not, perhaps, for sixth graders.

Many of our readers do not demonstrate through their intonation pattern, an understanding of this role for the colon. The sentence is read without a pause at the point of the colon, like a string of words. Many of the miscues on sinewy and stringy were non-word substitutions with high graphic similarity.

A similar sentence precedes Sentence 25 - Savage: wild not tamed, but resulted in fewer miscues. The intonation patterns suggest that perhaps sentence 26 was perceived as a continuation of the definition for Savage, or at least that readers didn't know where the syntactic pattern ended.

Although some sentences discussed seem to fit neatly into one category or another, it is most likely the case that most sentences with high miscues have several confounding features which result in high miscue frequencies.

Summary of Findings

Sentences resulting in highest MPWPR for each story were selected for analysis as an initial step in determining how and why miscues are more likely to occur in some places than others. From our initial evaluation of the data presented in Tables 1 and 2, we determined that miscue frequency was not simply a function of either sentence length or linguistic complexity as measured by the Schmidt-Kittel Linguistic Complexity Formula. Based on our theoretical model of the reading process, we investigated factors which might affect the reader's predictions of the written text.

We found that at least seven factors affect predictability and thus contribute to high miscue frequency:

1. Lack of prior contextual information.
2. Unfamiliar or unusual choice and use of lexical items.
3. Weak sentence structure.
4. Unpredictable but simple structures.
5. Unusual stylized syntax.
6. Complex syntactic structures.
7. A combination of any of the above.

For many sentences, the miscues themselves have a confounding effect in that once a miscue occurs in a sentence it is likely that others will follow. The reader will produce further miscues in an attempt to construct syntactically and semantically acceptable structures. In addition, sentences following those with high miscue rates will tend to have disproportionate numbers of miscues.

Discussion

Text difficulty has been a concern of educators for some time and has resulted in numerous "readability formulas" (Dale-Chall, 1948; Fry, 1968). Most of these formulas were designed for classroom use, with the goal of somehow matching the ability of the reader with the difficulty level of the text. Though matching author to reader may be an admirable goal, until recently we have lacked the theoretical base for analyzing text beyond superficial word, syllable, and sentence counts. Although some attempt was made to incorporate syntactic couplexity in some readability formulas (Botel and Granowsky, 1973), semantic and conceptual factors within connected discourse were more difficult to measure.

Within recent years, researchers have developed sophisticated tools for describing and analyzing the semantic structure of text (Kintsch, 1974; Frederiksen, 1975; Grimes, 1972). Using these and other similar research tools, studies on readers' comprehension of text through comparing the readers' recalls to the text have been conducted (Bridge, 1977; Marshall, 1976). Valuable insights into discourse comprehension, inference and representation of knowledge have emerged from such studies.

Kintsch (1977) has conducted research using propositional analysis aimed at discovering some factors adversely affecting text readability. He suggests the following factors: 1) proposition density, or the number of propositions relative to passage length; 2) constant introduction of new concepts as opposed to the repetition and development of a minimum number of concepts. This notion is supported by our research which revealed a high relative frequency of miscues for sentences in which a new, unpredictable, contextually inconsistent term

occurs (see previous discussion of Lack of Contextual Support and Unfamiliar or Unusual Lexical Items); 3) A lack of text coherence. The assumption here is that when the author does not explicitly represent relationships between various segments of the text, readers are forced to infer these relationships and supply the necessary linking information themselves. Kintsch suggests that this additional mental functioning may increase the processing load and slow the reading down. He points out that certain types of inferencing may effect readability more than others, and that further research will be needed to address this issue. Once again, our research lends some support to the validity of Kintsch's claim. Several of the sentences we studied required the reader to infer a relationship which had not been explicitly stated in the text. For instance, the reader is required to infer a causal relationship between sentence 22 and the previous context in Story 51. In order to comprehend sentence 76 in Story 44, it is necessary for the reader to infer a change in setting and characters. Certainly the metaphors in sentence 80, Story 53 and sentence 73 in Story 51 require complex inferences. All these sentences resulted in miscues for many of our readers, and some of these miscues indicated that the necessary inferences were not made; 4) the relative number of long term memory searches and reorganizations necessary in constructing the meaning for a text was cited as another possible factor.

Implications for Further Research

It seems clear that a synthesis of miscue analysis and text analysis is a promising means of discovering factors underlying text difficulty. Text analysis alone can provide a sophisticated semantic analysis of the text and the recall of the reader can contribute to our understanding comprehension. However, recalls of texts reveal only the product of comprehension, and in fact, this

product may be strongly influenced by factors such as the memory, selectivity, and self-confidence of the reader while retelling. Miscue analysis concerns itself with "on the spot" processing, or comprehending, and may, therefore, be better able to discover specific characteristics of text which prove difficult for several readers. In addition, while text analysis deals primarily with the semantic level of the text, miscue analysis also considers syntactic and morphological levels of text. Our research indicates that syntactic factors play an important role in miscue frequency. Perhaps particular relationships between propositions and their syntactic structures require more complex processing than others. Analyzing miscues in terms of the relationship between the syntactic and propositional structures of the text would be one way to explore this hypothesis. Furthermore, miscue analysis provides a way of studying the relationship between the comprehending process while reading and the overall comprehension expressed through the retellings.

Text difficulty can never be truly understood without investigating the interaction between readers and the text. As in any communication process, participants actively receive and furnish information. When a balance is reached between what each participant must give and take, successful communication is achieved. Perhaps "readability" is a function of the weight readers must bear in assuming their role in the communication process. Researchers now have more sophisticated, theoretically based tools with which to study both the writer's and the reader's contributions to written communication.

Future research in text difficulty and readability may not result in a fool-proof, easy-to-use readability formula, but it can contribute to a real understanding of the complex task of comprehending written language.

References

Allen, P. and D. Watson (Eds.). Findings of Research in Miscue Analysis: Classroom Implications, Urbana, Illinois: ERIC/NCTE, 1976.

Betts Basic Readers, Third Edition. E. A. Betts and C. M. Welch. New York: American Book Company, 1963.

Botel, M. J. and A. Granowsky. "A Syntactic Complexity Formula in Assessment Problems in Reading", W. H. MacGinitie (Ed.). Newark: International Reading Association, 1973. pp. 77-86.

Bridge, C. The text-based inferences generated by children in processing written discourse. Unpublished doctoral dissertation, University of Arizona, 1977.

Dale, E. and J. Chall. "A formula for predicting readability." Educational Research Bulletin, Ohio State University, 1948, 27, 11-20; 28, 37-54.

Frederiksen, C. H. "Representing Logical and Semantic Structure of Knowledge Acquired from Discourse." Cognitive Psychology, 1975, 7, 371-458.

Fry, E. "A readability formula that saves time." Journal of Reading, April, 1968, 11, 513-516.

Goodman, K. S. and Y. M. Goodman. Reading of American Children Whose Language is a Stable Rural Dialect of English or a Language other than English. NIE Final Report, NIE-C-00-3-0087, August, 1978.

Grimes, J. E. The Thread of Discourse. Ithaca, New York: Cornell University, 1972.

Kintsch, W. The Representation of Meaning in Memory. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Kintsch, W. and D. Vipond. Reading comprehension and readability in educational practice and psychological theory. Paper presented at Conference on Memory, University of Uppala, June 1977. In press in the proceedings of the conference, Lars-Goran Nilsson (ed.), Hillsdale, N. J.: Erlbaum Associates.

Marshall, N. "The structure of semantic memory for text." Unpublished doctoral dissertation. Ithaca, New York: Cornell University, 1976.

Schmidt, E. and J. Kittel. Schmidt-Kittel Linguistic Complexity Scale. Unpublished instrument, Seattle Pacific University, undated.